Target of this Project:
New York City Food venue Cluster locators, Popular food venues, Most populated food venues, popularity of various cuisines

Methodology

1. This project will help to understand the diversity of a neighbourhood by leveraging venue data from Four square's 'Places API' and 'k-means clustering' machine learning algorithm.
2. Exploratory Data Analysis (EDA) will help to discover further about the culture and diversity of the neighbourhood.
3. **Clients** would be the one who are interested to use this quantifiable analysis to understand the distribution of different cultures and cuisines over "the most diverse city on the planet - NYC".
4. Also, this project can be utilized by a new food vendor who is willing to open his or her restaurant. Or by a government authority to examine and study their city's culture diversity better.

Scope of the project

- This project will help to understand the diversity of a neighbourhood by leveraging venue data from Foursquare's ‘Places API’ and ‘k-means clustering’ machine learning algorithm.
- Exploratory Data Analysis (EDA) will help to discover further about the culture and diversity of the neighborhood.
- **Designated clients** would be the one who are interested to use this quantifiable analysis to understand the distribution of different cultures and cuisines over "the most diverse city on the planet - NYC".
- Also, this project can be utilized by a new food vendor who is willing to open his or her restaurant. Or by a government authority to examine and study their city's culture diversity better.

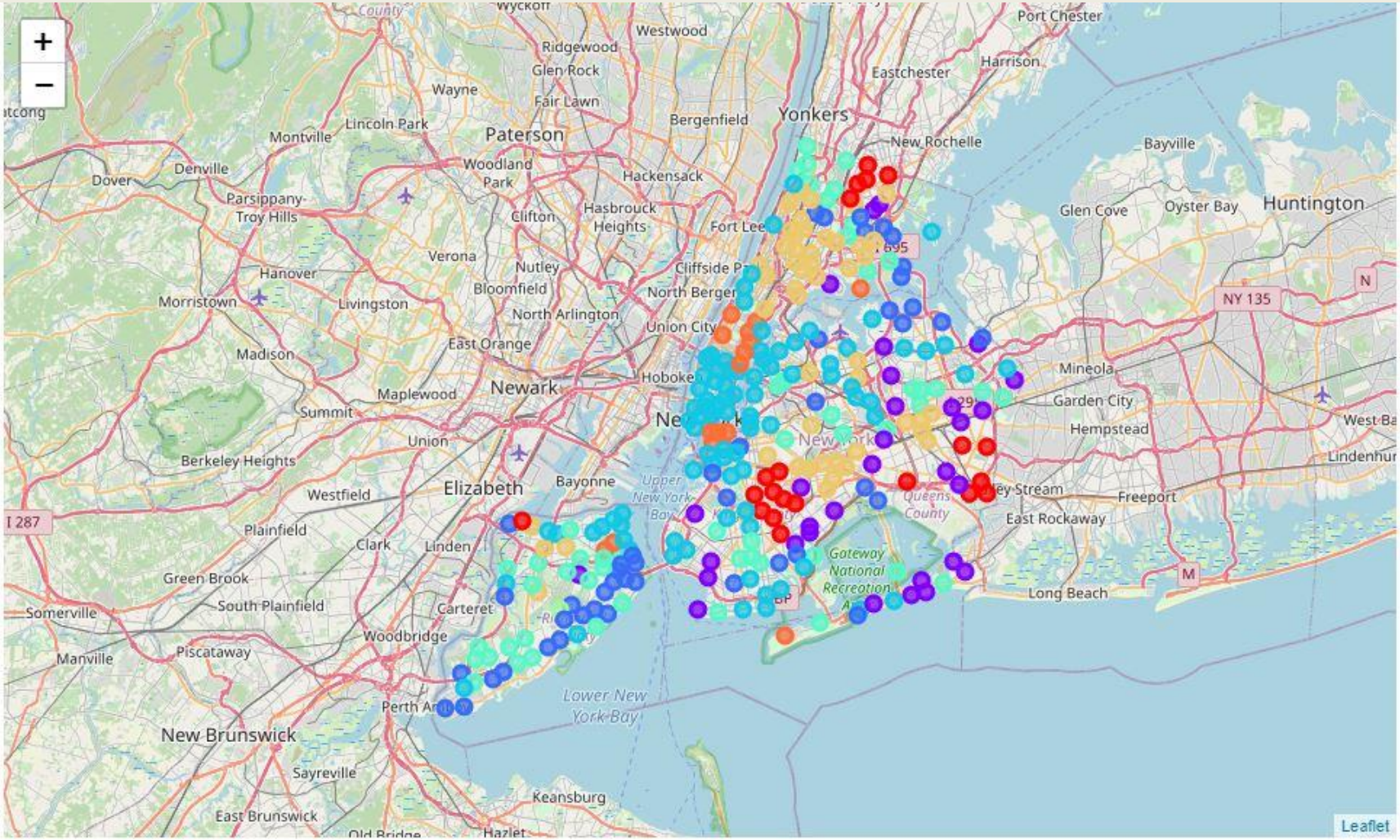
Data Acquisition and cleaning

- Cuisine preference, cuisine popularity, cuisine restaurants clusters.
- Information will be sourced from foursquare
- Link: https://geo.nyu.edu/catalog/nyu_2451_34572
- Link: https://developer.foursquare.com/docs
- Evaluate how many neighborhoods are there in NYC
- As, our aim is to segment the neighborhoods of NYC with respect to the *Food* in its vicinity. We need to proceed to fetch this data from all the 306 neighborhoods' venues.
- We find **14047 rows and 7 columns**
- We will need to categorize each venue in respect to the type of food. Eg: pizzeria, café, burger joint.
- Then we filter and separate venues based on the described food category
- We find out no of asian, italian, and other diversity's restaurants
- We visualize our datas at this point
- We analyse the clusters based on density of different cuisine's locality.

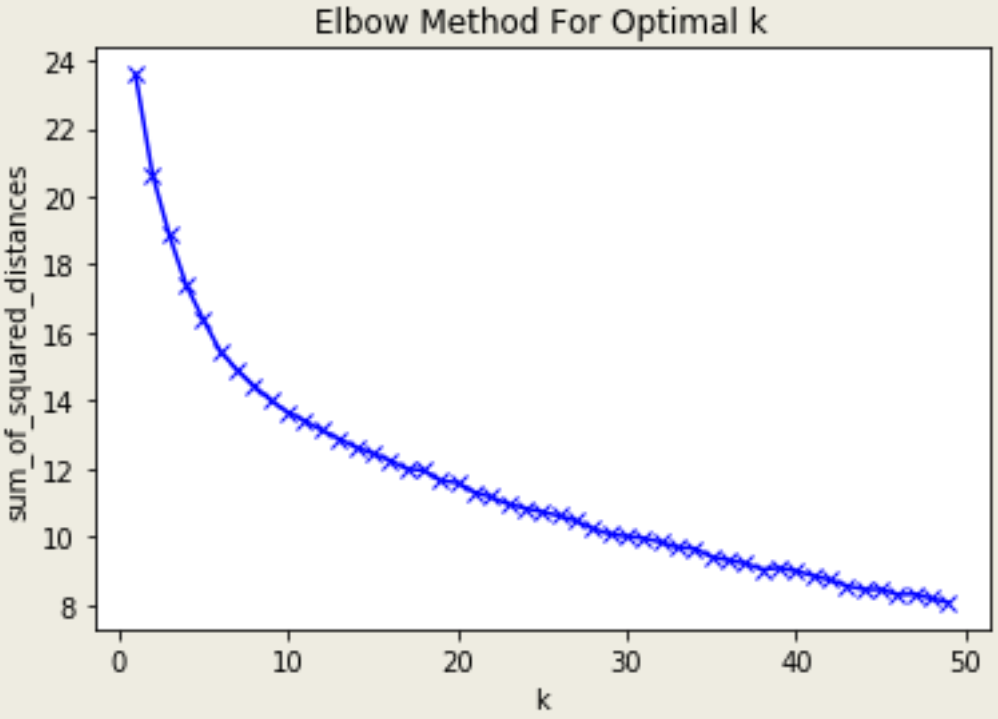
Libraries that we use are:

- Pandas
- Follium
- Numpy
- json
- Requests
- Geopy
- Matplotlib
- Sklearn.cluster
- Seaborn

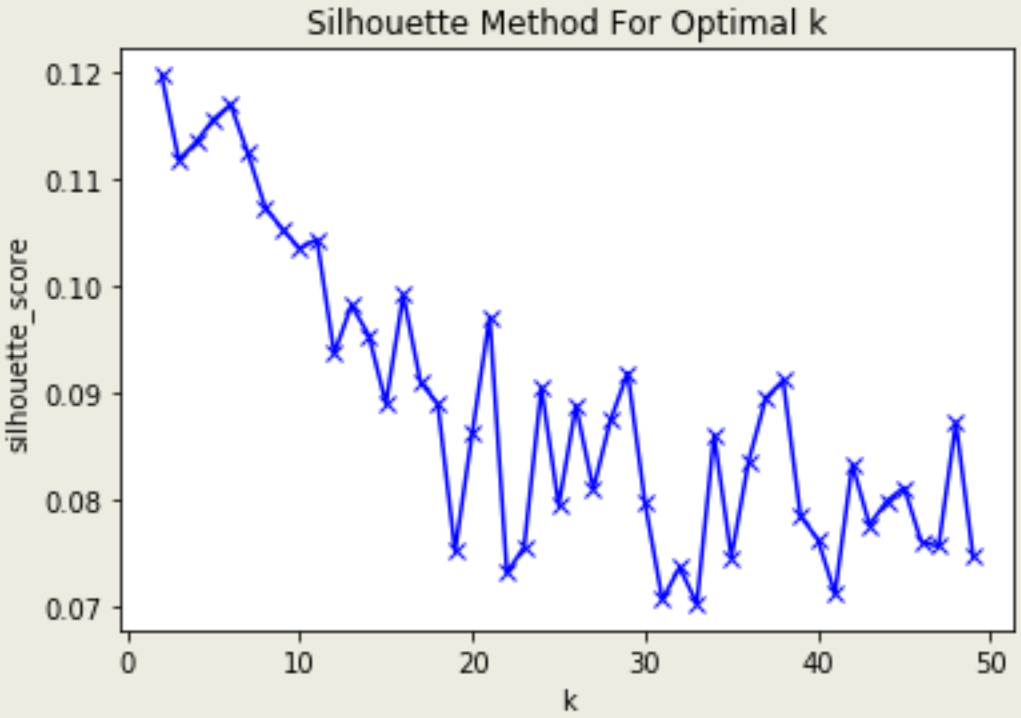
Follium map illustrating the Cluster cuisines in nyc
Color markers to show the diversity of cuisine



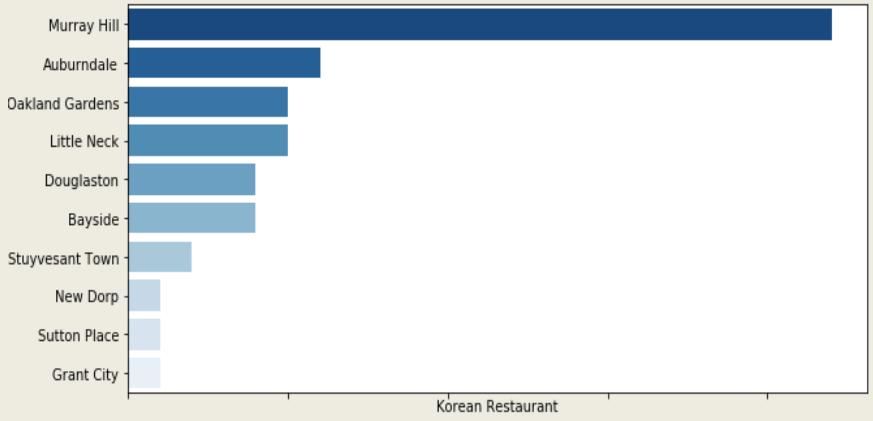
Analysis

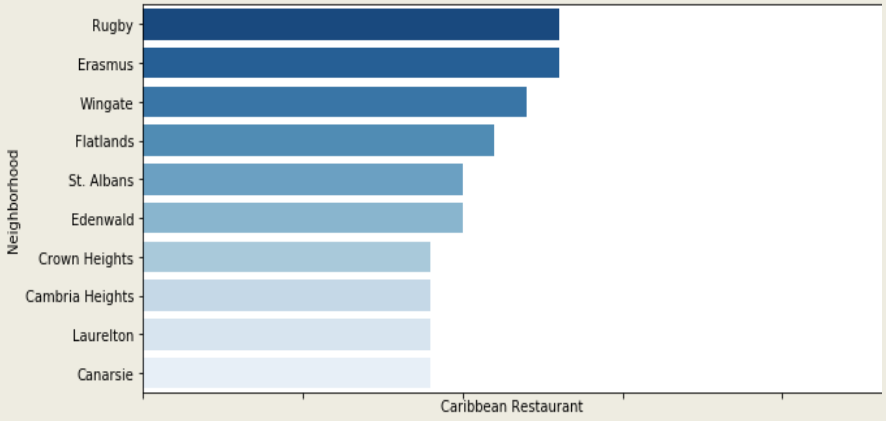

Elbow Method For Optimal k
sum_of_squared_distances
k
24
22
20
18
16
14
12
10
8
0
10
20
30
40
50

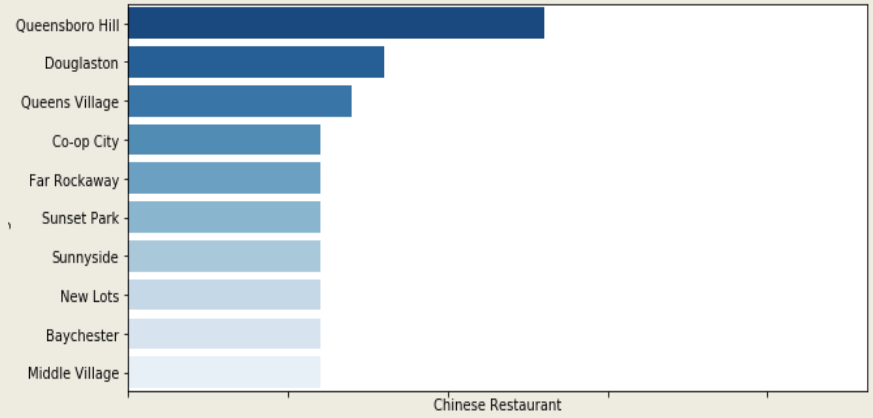
Analysis

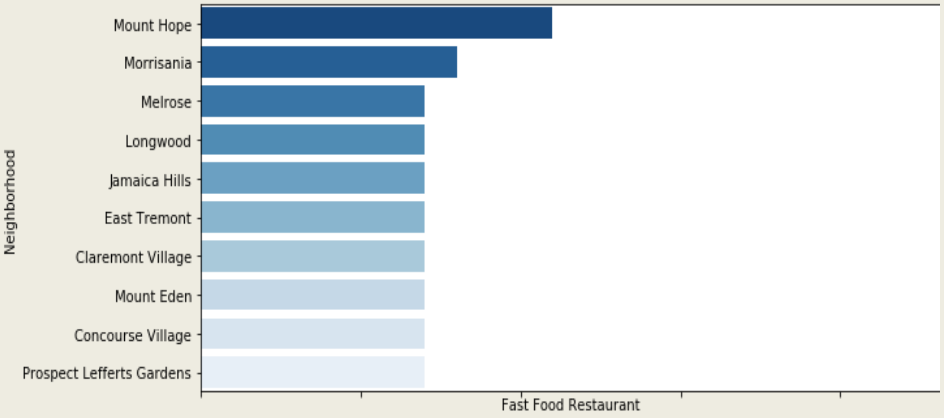

Silhouette Method For Optimal k
silhouette_score
0.12
0.11
0.10
0.09
0.08
0.07
0
10
20
30
40
50
k

Bar charts for x,y (cuisine, neighborhoods)

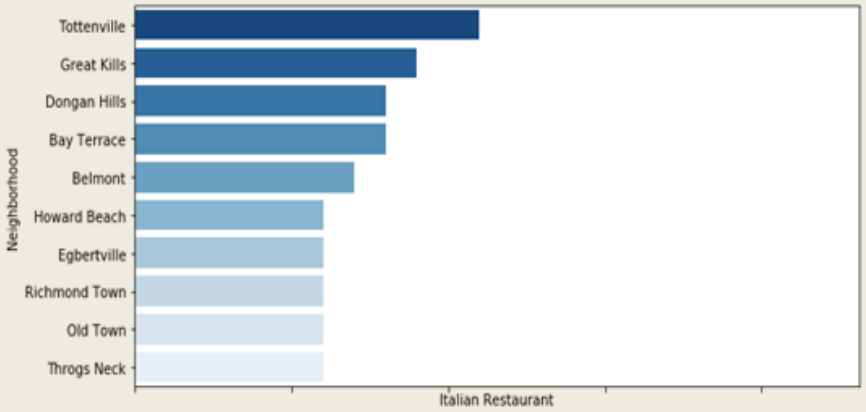


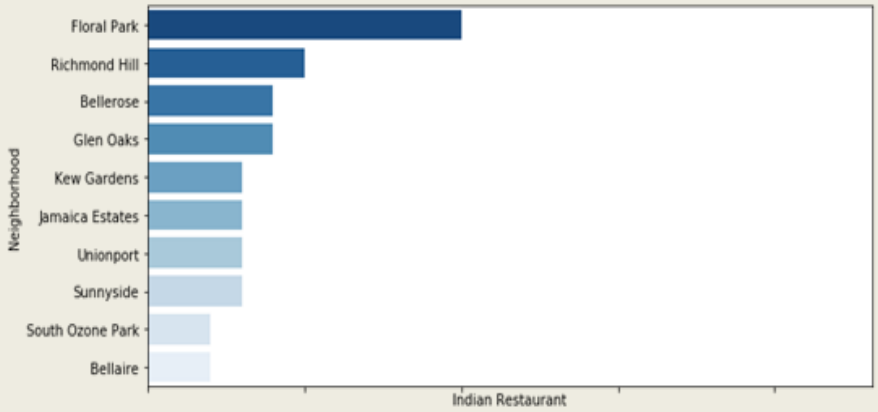


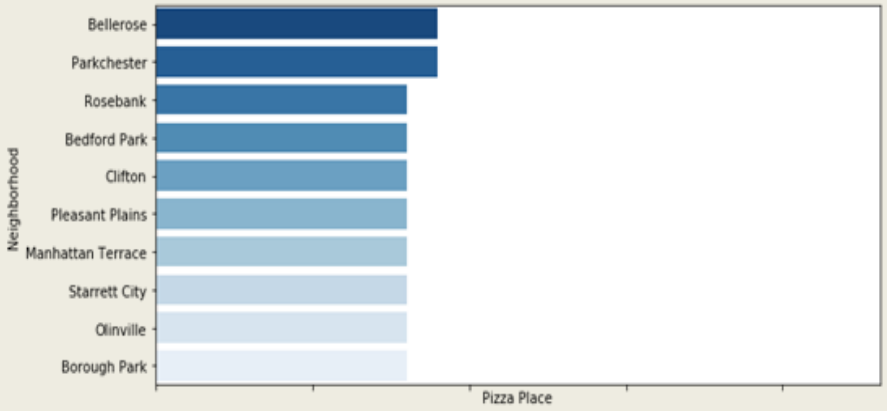


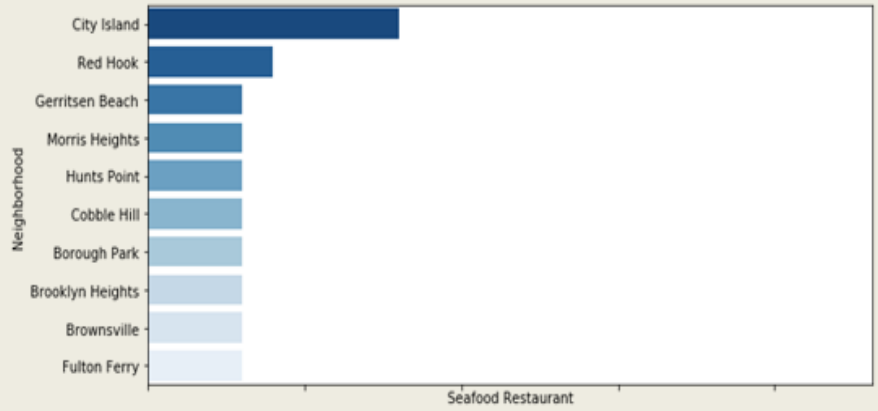


Bar charts for x,y (cuisine, neighborhoods)









This presentation will follow up with the final report_
Thank you.